REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 2002	Final Report (February 1999 - March 2002)

4. TITLE AND SUBTITLE
The Effects of Augmented Physiological Feedback on Detection of Deception

5. FUNDING NUMBERS
N00014-99-1083

6. AUTHOR(S)
Robert M. Stern, Ph.D.; John C. Kircher, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DoD Polygraph Institute
7540 Pickens Avenue
Fort Jackson, SC 29207

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
DoDPI99-P-0004/DoDPI02-R-0009

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Public release, distribution unlimited.

12b. DISTRIBUTION CODE
A

13. ABSTRACT *(Maximum 200 words)*
The purpose of this study was to assess the effects of two types of augmented physiological feedback (APF) on the reliability and accuracy of probable-lie comparison question test (CQT). Two hundred and ten college students participated in the study, half of whom were guilty of a mock crime and half innocent. During questioning, one group received skin conductance feedback, a second group received composite feedback (skin conductance, cardiograph, and respiration), and a third group received no feedback. The results indicated that APF did not increase detection rates above that of no-feedback group in this study. However, APF did decrease the rate of habituation during repetition of the question sequences thus allowing for greater discrimination between innocent and guilty participants as the CQT progressed.

14. SUBJECT TERMS

15. NUMBER OF PAGES
42

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

DoDPI02-R-0009

The Effects of Augmented Physiological Feedback on Detection of Deception

Robert M. Stern, Ph.D. The Pennsylvania State University, John C. Kircher, Ph.D.
University of Utah

March 26, 2002

Department of Defense Polygraph Institute
7540 Pickens Avenue
Fort Jackson, South Carolina 29207

THE EFECTS OF AUGMENTED PHYSIOLOGICAL FEEDBACK ON DETECTION OF DECEPTION

Final Report of ONR Grant N00014-99-1-1083

January 2, 2002 (Revised March 26, 2992)

Principal Investigator:

Robert M. Stern, Ph.D.
Distinguished Professor of Psychology
512 Moore Building
The Pennsylvania State University
University Park, PA 16802
Phone: 814-865-1712
Fax: 814-863-7002
E-mail: RS3@PSU.EDU

Co-Investigator:

John C. Kircher, Ph.D.
Professor of Educational Psychology
University of Utah
Salt Lake City, UT 84112
Phone: 810-581-7130
E-mail: KIRCHER@GSE.UTAH.EDU

ABSTRACT

The purpose of this study was to assess the effects of two types of augmented physiological feedback (APF) on the reliability and accuracy of probable-lie comparison question tests (CQT). Two hundred and ten college students participated in the study, half of whom were guilty of a mock crime and half innocent. During questioning, one group received skin conductance feedback, a second group received composite feedback (skin conductance, cardiograph, and respiration), and a third group received no feedback. The results indicated that APF did not increase detection rates above that of the no-feedback group in this study. However, APF did decrease the rate of habituation during repetition of the question sequences thus allowing for greater discrimination between innocent and guilty participants as the CQT progressed.

This research was funded by the Department of Defense Polygraph Institute, Fort Jackson, South Carolina as project DoDPI99-P-0004. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of Defense or the U. S. Government.

THE EFFECTS OF AUGMENTED PHYSIOLOGICAL FEEDBACK ON DETECTION OF DECEPTION

The purpose of this study was to assess the relative effects of two types of physiological feedback on the reliability and accuracy of probable-lie comparison question tests (CQT). The CQT is the most widely used method of psychophysiological detection of deception by field polygraph examiners (Ben-Shakhar, 1991). Decisions about a suspect's involvement in criminal activities are based upon within-participant comparisons of physiological reactions to questions relevant to the criminal investigation (e.g., theft of a determined amount of money) and to probable-lie comparison questions. Comparison questions address a general content area that is related to, but excludes the specific criminal activity in question (Reid & Inbau, 1977). For example, if the criminal activity under investigation pertained to the theft of money, a comparison question might be, "Before the age of 21, did you ever take something that didn't belong to you?" Comparison questions are intentionally vague and are nearly impossible to answer truthfully with an unqualified "No". During a pretest interview, suspects are embarrassed or intimidated into answering "No". If an affirmative response is given to a probable-lie question, the question is reworded so that suspect will ultimately answer in the negative, which is probably a lie.

The CQT assumes that the suspect's degree of involvement with each type of question and the relative amount of concern that each type of question would evoke is diagnostic (Stern, Breen, Watanabe, & Perry, 1981). For guilty suspects,

polygraph procedures are designed to reinforce their concern that deceptive answers (i.e., a "No" response) to crime relevant questions will be detected. Even though guilty suspects would answer "No" to probable-lie comparison questions, the crime-relevant questions are expected to cause more concern since relevant questions deal specifically with the matter under investigation. For innocent suspects, only the probable-lie comparison questions are answered deceptively. Because innocent suspects had no involvement in the criminal activity in question, the probable-lie questions are expected to elicit greater concern about being (in their opinion, falsely) detected than the crime-relevant questions. If these goals are achieved, guilty suspects should show stronger physiological responses to the relevant questions than to the probable-lie questions, whereas innocent suspects should show stronger reactions to the probable-lie questions.

Although these predictions have been confirmed in laboratory and field settings (Office of Technology Assessment, 1983; Raskin, Honts, & Kircher, 1997), it is equally important to understand the psychophysiological processes that underlie these findings and to devise techniques that would exploit these processes in order to increase detection accuracy rates.

Kircher (1981) offered a theoretical framework to explain the differential reactivity to probable-lie and relevant questions seen in CQTs. He suggested that when a suspect intends to answer a question deceptively during a polygraph test, the presentation of the question signals the occurrence of an involuntary physiological reaction. The participants' expectation that their bodies will reveal deception with large physiological reactions is established and reinforced during the pretest phase of the polygraph examination. According to this view, the participant's expectation that a large involuntary reaction will accompany deception is an essential component of a valid polygraph outcome.

If participants expect large autonomic reactions when they lie, it is reasonable to assume that when they lie, they will attempt to monitor and suppress these internal changes. Borrowing from Kahneman's (1973) theory of attention and effort, Kircher (1981) hypothesized that mental effort is required to suppress and monitor the leakage of incriminating information. That is, the participant must mobilize and expend energy to perform these cognitive activities. According to this hypothesis, the physiological changes recorded by the polygraph are indicators of the amount of mental effort or attention required by participants to monitor and suppress their autonomic responses to test questions. This hypothesis predicts that the perception of increased autonomic reactions will create a positive feedback loop that requires additional mental effort and prolongs the participant's cognitive appraisal of yet another threatening event. The perception of physiological arousal that occurs after the presentation of the test question may explain observed increases in the duration of physiological responses associated with deception (e.g., Kircher & Raskin, 1988; Raskin et al., 1988).

The theory also predicts that the proposed use of augmented physiological feedback (APF) will increase the differences between physiological reactions to comparison and relevant questions and thereby improve discrimination between truthful and deceptive participants. When the participant is deceptive, feedback that a strong autonomic response has occurred would be viewed as an aversive event. Like the test question that initiated the response, the feedback would threaten the participant. If the feedback is public, such that the participant knows the polygraph examiner is also hearing it, it should increase the threat to the participant, or in other words, it should signal to the participant that he/she is revealing him/herself.

Previous evidence suggests that guilty and innocent suspects respond differentially to probable-lie and crime-relevant questions (e.g., OTA, 1983). Guilty suspects react more strongly to crime-relevant questions than to comparison questions, whereas innocent suspects react more strongly to comparison questions than to crime-relevant questions. It is hypothesized that if APF increases the amount of involvement or attention allotted to questions that already pose the greatest threat to the suspect, then guilty suspects should appear more deceptive on their polygraph tests by showing an even greater response to crime-relevant questions, whereas innocent suspects should appear more truthful by showing a greater response to probable-lie questions versus crime-relevant questions. Hence, it should be easier to distinguish between truthful and deceptive suspects, thereby increasing the accuracy of the CQT.

Using the Guilty Knowledge Test, Stern, Breen, Watanabe, and Perry (1981) tested for the hypothesized beneficial effects of APF on detection rates. Participants in the APF condition received an auditory signal that varied with changes in their heart rate or skin resistance (SR), whereas control participants received no feedback. All participants were given two GKT polygraph tests: the first test was based on a geometric figure chosen by the participant from a list of five (low personal relevance test), and the second test concerned the participant's Social Security Number (SS#) (high personal relevance) that was embedded among a list of four other SS#s. Participants answered "No" to each of the four alternatives for both tests. Stern et al. found that discrimination between critical and noncritical items, based on participants' SR responses, was statistically significantly greater for the SR feedback group than the no feedback group. An effect for personal relevance was also found, such that accuracy for tests concerning SS# was statistically significantly greater than the accuracy for tests about the geometric figures.

The second experiment reported by Stern et al. (1981) assessed the effects of APF on innocent and guilty participants involved in a simulated murder plot. Participants in the guilty condition studied a document that contained several details about their role in the murder plot. Innocent participants studied a document that contained the same details, but the details were totally unrelated to any murderous activity. Half of each group was assigned to a SR Feedback condition, and the remaining participants served as a No-Feedback control. Although no statistically significant effect of feedback was found participant mean SR response to critical items was greater in the feedback condition than the no-feedback condition for both guilty and innocent participants. The lack of statistical significance is probably due to a ceiling effect seen in the No-Feedback Group, such that detection rates in this control condition were high enough that any added benefit of an experimental procedure would be very difficult to detect without a very large sample size.

The results of the Stern et al. (1981) experiments are consistent with the prediction that feedback will enhance physiological responses to items of greater relative importance to the suspect. If this hypothesis is correct, then APF should differentially affect the responses of guilty and innocent suspects to relevant and probable-lie questions, respectively, using the CQT. The present experiment was designed to test that prediction.

Another investigation of the effects of auditory biofeedback on the Guilty Knowledge Test was conducted by Timm (1987). He found that electrodermal feedback statistically significantly enhanced detection efficiency associated with respiration amplitude changes, but that skin conductance detection efficiency was not statistically significantly affected. These results were similar to the results found in the Stern mock murder experiment. The null results found in the studies by Stern, et al. and Timm may be due to a ceiling effect for the No Feedback condition, as they suggest; however, the null results might also be due to low power, as the Stern et al. study employed only 52 participants, and the Timm study employed 68 participants. In the present study, the sample size was increased to 210 participants, which provided an 80% probability of detecting moderate (i.e., .4-.6) differences between feedback conditions.

In addition to the techniques employed by Stern et al. (1981) and Timm (1987), this study explored alternative methods of providing feedback to participants, as well as alternative analysis procedures used to identify and classify innocent or guilty suspects. Specifically, the study also addressed the question of whether or not feedback based on electrodermal activity alone results in a more

reliable index of guilt than a composite of several physiological measures. Stern et al. (1981) had greater success with electrodermal feedback than heart rate feedback. However, with the current state of computer technology, a composite index of arousal based on electrodermal, cardiovascular, and respiration measures may be generated and presented to the participant in real-time. Since some examinees show little or no electrodermal activity or their electrodermal responses quickly habituate (defined as a decrease in amplitude as a result of repeated exposure to the polygraph questions), the use of a composite index should allow investigators to provide those individuals with variable feedback even in the absence of changes in electrodermal activity.

Stern et al. (1981) classified participants as truthful or deceptive and assessed the number of correct hits and correct rejections. In addition to reporting decision accuracy, the present study tested for effects of guilt and APF on discrete measures of electrodermal, cardiovascular, and respiratory activity.

Method

Participants

Two hundred-ten college students (males =71; age range 18-60) from the Pennsylvania State University volunteered to participate in this study. Participants were in good health, free of psychotropic medication and had not previously taken a polygraph test. Participants received extra credit for their psychology courses; and, if found innocent on the polygraph test, they were given $20.

The participants were randomly assigned to one of six conditions in a balanced 2 X 3 between-groups factorial design. Specifically, there were two levels of Guilt (guilty and innocent) and three levels of Feedback (no feedback, skin conductance, and composite. The university's Institutional Review Board approved the study protocol and informed consent document prior to participant recruitment.

The polygraph examiner was a graduate student who had been trained to use standard polygraph procedures; the examiner did not have any previous academic interactions with any of the participants.

Apparatus

Physiological Data Collection: The CPSLAB system (Scientific Assessment Technologies, SLC, UT) was used to configure the data collection hardware, specify storage rates for the physiological signals, and build automated data collection protocols. CPSLAB was also used to collect, edit, and score the physiological data.

The physiological data acquisition subsystem (PDAS) of CPSLAB generated analog signals for thoracic and abdominal respiration, skin conductance, finger pulse amplitude, and EKG. Each of the five analog signals was digitized at 1000 Hz with a Metrabyte DAS 16F analog-to-digital converted installed in the CPSLAB computer. The CPSLAB computer collected and stored the polygraph charts.

Respiration was recorded from two Hg strain gauges secured with Velcro straps around the upper chest and the abdomen just below the rib cage. Resistance changes were recorded DC-coupled with a 2-pole, low-pass filter, fc = 13Hz.

Skin conductance was obtained by applying a constant voltage of .5V to two UFI 10mm Ag-AgCl electrodes filled with .075M NaCl in a Unibase medium. The electrodes were strapped with adhesive to the middle phalanx of the fourth and fifth fingers of the left hand. The signal was recorded DC-coupled with a 2-pole, low-pass filter, fc = 6 Hz.

Finger pulse amplitude was obtained from a UFI photoplethysmograph attached to the index finger of the left hand with a Velcro strap. The signal from the photocell was AC-coupled with a 0.2-second time constant and a 2-pole, low-pass, fc = 10 Hz.

The electrocardiogram was obtained from the limb Lead II configuration of Einthoven's Triangle using disposable, pre-gelled Ag-AgCl snap electrodes. The PDAS generated a 20 ms square wave pulse that coincided with the R-wave in the electrocardiogram. The square wave from the PDAS was routed to the analog-to-digital converter, and the CPSLAB software measured and stored the time between successive pulses (interbeat intervals).

The 1000 Hz samples for each channel were reduced prior to storing them on the hard disk by computing the mean of samples for successive data points. Respiration and electrodermal channels were stored at 10 Hz. Cardiograph and finger pulse signals were stored at 100 Hz. The cardiotachometer produced an interbeat interval measured to the nearest ms for each heartbeat.

Feedback. The analog respiration, skin conductance, and cardiograph signals along with event marks were routed to a second computer equipped with a Metrabyte DAS 8 analog-to-digital converter. The second computer was programmed to provide the appropriate type of feedback (if any) to the participant. Auditory feedback was produced by the speaker in the PC. The auditory feedback was a tone that varied in pitch with changes in skin conductance or the composite index of arousal. The composite feedback was based on a weighted combination of changes in skin conductance (50%), cardiograph (25%), and respiration (25%). The feedback began at question onset, lasted for 20 seconds, and then was turned off until the next question was presented.

Procedure. Prospective participants registered on the Internet for an appointment to participate in the experiment. When the participant arrived, an envelope addressed to the participant was taped to the door of the meeting room. Instructions within the envelope directed the participant to enter the room, close the door, read and sign an informed consent form, complete the polygraph accuracy questionnaires, and then play a cassette recorder that presented instructions over earphones.

Guilty participants received tape-recorded instructions to commit a mock theft of $20 from a purse in a desk drawer in the room where they received their instructions. The participant searched the desk for the purse, took the $20, concealed it on his/her person, and then reported to the laboratory where the polygraph test was administered. Innocent participants listened to a general description of the crime and then reported to the laboratory for their polygraph examination.

All participants were told that a polygraph expert who didn't know if they had committed the theft would give them a polygraph test. They were told that the examiner would use a computer to assist in the analysis of their polygraph charts. They were also told that they would receive course credit and would be paid $20 if they passed the polygraph test; but they would receive only the course credit and not be paid if they failed the test or the test was inconclusive. Thus, innocent participants were paid the bonus if they were correctly classified by the computer as innocent, whereas guilty participants were paid $20 if the computer incorrectly classified them as innocent. Throughout the polygraph procedure, the polygraph examiners remained blinded to the participant's condition. Once the computer decision was revealed to the participant, the polygraph examiners received documentation from an assistant who assigned the condition to the participant. The assistant did not participant in any aspect of the polygraph test, other than determining the participant's guilt and feedback status.

Pretest. When the participant arrived at the laboratory, the polygraph examiner introduced herself, obtained some demographical data, and reviewed the test questions with the participant. Standard field polygraph procedures were used. Relevant questions that pertain to the theft and the sacrifice relevant were reviewed first, probable lie questions were reviewed next, and the neutral questions were reviewed last. The test questions were as follows:

Test Questions

(Sacrifice Relevant)	1. Do you intend to answer truthfully all of the questions about the theft of the $20?
(Neutral)	2. Do you live in the United States?
(Probable lie)	3. Before the age of _____, did you ever take something that didn't belong to you?
(Relevant)	4. Did you take that $20 from the purse?

(Neutral) 5. Is today _____?

(Probable lie) 6. During the first _____ years of your life, did you ever do anything that was dishonest or illegal?

(Relevant) 7. Did you take that $20?

(Neutral) 8. Is your first name _____?

(Probable lie) 9. Between the ages of _____ and _____, did you ever lie to get out of trouble?

(Relevant) 10. Do you have that $20 with you now?

After reviewing the test questions, sensors were attached to the participant. The polygraph examiner then described the role of the autonomic nervous system in the detection of deception and administered a standard numbers test. Consistent with field practice, participants were informed that they showed their strongest reaction when they lied about the number they chose and showed smaller reactions when they were truthful.

No APF was given during the numbers test. Since the numbers test is a relatively weak manipulation, a high percentage of participants did not actually show their strongest reaction to the chosen number. If participants were to receive APF that revealed that they showed a relative weak reaction to the chosen number, it would not be possible for the polygraph examiner to claim that they did. Moreover, if participants learned from the APF that the technique failed to detect their deception during the numbers test, the accuracy of the subsequent CQT might suffer (Bradley & Janisse, 1981).

Following the numbers test, participants in the APF conditions were informed about the nature of the feedback they would receive during the CQT. Participants in the skin conductance and the composite feedback condition were told that a tone would be presented during the polygraph test. They were told that this tone would rise in pitch as a function of the magnitude of their physiological response to each question.

Interrogation. The probable-lie test was then administered. The question sequence was presented five times, and the interval between repetitions of the question sequence was from one to three minutes. The order of neutral and probable-lie questions varied over repetitions of the question sequence such that each neutral and each probable-lie question at least once preceded each relevant question. The interval between question onsets was a minimum of 35 s.

At the conclusion of the test, the sensors were removed, and the participant was asked to complete the post-test questionnaire. The probability that the participant answered the relevant questions truthfully was then computed using the CPS algorithms developed at the University of Utah (Kirchner & Raskin, 1988; 1994). According to the CPS algorithm, if the probability that the participant was truthful exceeded 0.70, the participant was classified as innocent and was awarded the $20 and course credit. Otherwise, the participant was given only course credit. The participant was then debriefed and informed that the study was finished.

Data Analysis

Dependent Variables

Dependent measures consisted of computer measurements and computer decisions. The CPSLAB software provided the computer measurements and the CPS program provided computer decisions.

Computer Measurements. The CPSLAB software measured skin conductance amplitude (SC amplitude), cardiograph amplitude, and respiration excursions as follows:

SC Amplitude. A SC response curve was defined as the series of samples taken at 10 Hz from question onset to the 20th post-stimulus second. The computer identified points of inflection in the response curve and measured the difference between each low point and every succeeding high point. SC Amplitude was quantified as the greatest observed difference between a low and high point.

Cardiograph Amplitude. CPSLAB identified the time and level of each systolic point in the cardiograph. The systolic points were used to create a second-by-second systolic curve from question onset to 20 seconds post-question onset. Another second-by-second curve was computed from the diastolic points. The mean of the systolic and diastolic points was then compared for each second. Cardiograph amplitude was extracted from the mean response curve in the manner described above for SC amplitude.

Respiration Excursion. Excursion was operationalized as the sum of absolute values of differences between successive 10 Hz samples of respiration obtained from question onset to 20 seconds post stimulus (100 discrete samples). A separate sum of absolute values (excursion) was obtained for thoracic and abdominal respiration. The mean of thoracic and abdominal excursion was computed for each test question. The repeated measurements of thoracic and abdominal respiration excursions, taken separately, were transformed to standard scores. Respiration excursion was defined as the mean of the standard measurements of thoracic and abdominal excursions.

For each physiological measure, an index of differential reactivity to relevant and comparison questions were computed in the manner described by Kircher and Raskin (1988). Briefly, the three probable-lie and three relevant questions on each of the first three charts provided 18 repeated measures of a physiological component. The 18 measurements for a physiological measure (e.g., SC amplitude) were converted to standard scores.

Mean standard scores for relevant questions were subtracted from mean standard scores for comparison questions. The sign of the computer index indicated which question produced the stronger reaction, and the magnitude of the score provided a precise measure of the difference between reactions to the two types of questions.

For all measures except respiratory excursion, a large measured response was indicative of a strong physiological response to a question. However, relatively small measured responses for respiration indicated greater respiratory suppression, which is associated with deception (Kircher & Raskin, 1988; Timm, 1982). Therefore, the sign of the standardized scores for respiration was reversed so that higher scores indicated stronger reactions, consistent with the other physiological measures.

Computer Decisions

The procedures used for making computer decisions paralleled those used by field polygraph examiners who perform numerical evaluations of polygraph charts. If the computer analysis of the first three charts yielded a probability of truthfulness of .70 or greater, or .30 or less, the participant was classified as innocent or guilty, respectively. If the computer analysis was inconclusive after three charts, the final two charts were included in the computer analysis. Participants were classified as inconclusive only if after five charts, their probability of truthfulness remained between .30 and .70.

Reliability of Computer Measurements. Coefficient alpha assessed the internal consistency of computer indices of differential reactivity. To compute coefficient alpha, an index of differential reactivity was computed for each of the 15 comparison-relevant question pairs obtained from the five charts. The 15 difference scores were treated as responses to 15 items on a test (Kircher & Raskin, 1988). If APF captured the attention of participants and reduced random variation in how they processed test questions, the reliability of physiological measures should be greater for participants who received APF than for those in the no-feedback control conditions.

Statistical Tests and Power. A series of univariate comparisons were performed to determine if there were statistically significant effects for Guilt, Feedback, and Sex on each computer index of differential reactivity. For the proposed analyses, the power to detect a medium effect (0.5 within-group standard deviation) exceeded .90 with 210 participants; hence this design had sufficient power to determine if feedback was statistically significantly better or worse than no-feedback. Planned Guilt X Type of Feedback interaction contrasts (Keppel, 1991) were performed to determine if APF affected discrimination between guilty and innocent participants.

Based on the results of those statistical tests, the type of APF that maximized discrimination between truthful and deceptive participants was identified; and the data from that condition were used to test if APF reduced habituation of physiological responses to comparison and relevant questions. A Guilt X Feedback X Charts split-plot ANOVA was performed for each index of differential reactivity. Feedback had two levels (the selected APF and No Feedback); and Charts consisted of a repeated measure with three levels. A more rapid decline in the discrimination between guilty and innocent participants across charts was expected in the no-feedback condition. The selected APF condition was expected to reduce habituation and improve discrimination between guilty and innocent participants across charts.

Analyses of Computer Decisions. Yate's corrected chi-square tests were used to test if there existed differences in accuracy between feedback and no-feedback conditions and between types of feedback. For each of these analyses, a dichotomous decision rule ensured that all participants were classified as truthful or deceptive. "Truthfulness" was defined as having a probability of .50 or higher. These chi-square analyses were performed separately for guilty and innocent participants.

Analyses of Physiological Waveforms. **To explore the possibility that APF affects the duration of a physiological response, rather than its amplitude, traditional split-plot ANOVA was used to test for differences in shapes of physiological responses obtained for comparison and relevant questions over the 20-second interval that followed question onset (Kircher, Woltz, Bell & Bernhardt, 1998; Podlesny & Raskin, 1978). These analyses included second-by-second skin conductance, cardiograph, respiration, finger pulse amplitude, and heart rate measures. The between-groups factors consisted of Guilt (2 levels), Feedback (3 levels), and Sex (2 levels); within-participants factors consisted of Question Type (Comparison and Relevant), Charts, and Time (20 seconds). Vagal tone, was measured via the Porges-Bohrer algorithm every five seconds during the 20 seconds that followed question onset. Therefore, the time factor in the ANOVA for vagal tone had four levels rather than 20.**

Results

Missing Values

Forty-eight of the 1050 charts for the 210 participants (210 X 5) were missing due to participants' reports of fatigue (~20) or due to data collection malfunction (~20); but there was no statistically significant relationship between the loss of charts and group assignment. The first three charts were available for every participant but one. That participant's missing first chart was replaced with her second chart. Charts 4 and 5 were used only to make decisions and only in the event that the outcome based on the analysis of the first three charts was inconclusive. In

two cases, the fourth and fifth charts were unavailable for participants with inconclusive outcomes after the first three charts. For those participants, the test was considered inconclusive.

Computer Decisions and Reliability

Table 1 presents the percentage of cases classified correctly, incorrectly, and as inconclusive. Table 1 also presents the percent correct decisions including inconclusive outcomes for each group of participants. Table 2 shows the reliability of differential reactivity measured across the 15 probable-lie/relevant question pairs in the five repetitions of the question sequence (charts). Mean reliability as measured by coefficient alpha was slightly higher for the APF groups than for the control group.

Effects of Gender on Dependent Measures

Preliminary Guilt X Feedback X Gender ANOVAs revealed no main or interaction effects on SC, cardiograph, or respiration measures that involved Gender. Therefore, Gender was dropped as a factor from all subsequent analyses.

Effects of SC Feedback on Physiological Measures

To determine if continuous auditory feedback of SC activity increased discrimination between guilty and innocent participants, a separate Guilt X Feedback interaction comparison was performed for each of the three computer

indices of differential reactivity. Guilt had two levels (guilty and innocent) and Feedback had two levels (no-feedback and SC-feedback). The means for SC amplitude, cardiograph amplitude, and respiration excursion are plotted in Figure 1.

Figure 1 suggests that discrimination between guilty and innocent participants tended to be greater in the SC-feedback condition than in the no-feedback condition for measures of SC amplitude and cardiograph amplitude and less for respiration excursion. However, the interaction comparisons for SC amplitude, $t(204) = 1.51$, $p < .14$, cardiograph amplitude, $t(204) = 1.09$, and respiration excursion, $t(204) = -1.31$, were not statistically significant.

Effects of Composite Feedback on Physiological Measures

To determine if composite feedback improved discrimination between guilty and innocent participants, separate Guilt X Feedback interaction comparisons of no-feedback and composite feedback conditions were performed for SC amplitude, cardiograph amplitude, and respiration measures. The means for the three physiological measures are plotted in Figure 2. Again, none of the interaction comparisons was statistically significant.

Effects of Feedback on Dichotomous Computer Decisions

Table 3 presents the percentage of cases classified correctly and incorrectly when participants were considered truthful if the probability of truthfulness exceeded 0.50 and were considered deceptive if the probability of truthfulness was less than .50. Consistent with the results reported above for individual physiological measures, chi-square analyses revealed no statistically significant differences between no-feedback and SC feedback conditions, between no-feedback and composite-feedback conditions, or between the SC-feedback and composite-feedback for either innocent or guilty groups.

Effects of Feedback on Habituation Rates

Guilt X Feedback X Charts split-plot ANOVAs were conducted to test the prediction that APF would reduce the habituation of SC, cardiograph, and respiration responses to repeated presentations of comparison and relevant questions. One ANOVA was performed to compare the no-feedback and SC-feedback conditions, and another ANOVA compared the no-feedback and composite-feedback conditions. The first set of analyses, displayed graphically in Figure 3, was limited to the first three polygraph charts. This three-chart analysis was conducted independent of the full five-chart analysis to determine APF effects on habituation in a situation more similar to a field polygraph test, where only three charts are collected. P-values based on Huynd-Feldt corrected degrees of freedom were used to decide if an effect was statistically significant. Results suggest the Guilt X Feedback X Charts interaction effect on SC amplitude was statistically significant when participants who received SC feedback were compared to those who received no feedback, $\underline{F}(2, 272) = 6.84$, $\underline{p} < .01$, $\eta^2 = .05$. The means for the first three charts are presented in the left panel of Figure 3. Figure 3 reveals a clear difference between guilty and innocent groups on the first two charts whether or not the participants received auditory SC feedback. However, on the third chart, discrimination between guilty and innocent participants was statistically significantly greater for participants who received APF than no-feedback. A similar effect on SC amplitude emerged when participants who received composite feedback were compared to those who received no feedback, $\underline{F}(2, 272) = 3.70$, $\underline{p} < .05$, $\eta^2 = .03$. The means for the no-feedback and composite feedback conditions are presented in the center panel of Figure 3. Similar to the SC feedback condition, discrimination between guilty and innocent participants by the third chart was greater for participants who received composite APF than for those who received no feedback.

To determine whether or not the effects of APF persisted in further chart presentation, a second analyses which included the fourth and fifth charts was conducted. The Guilt X Feedback X Charts effect was still significant for the No-Feedback vs SC Feedback comparison, $p < .02$ with Geisser-Greenhouse corrected df. As predicted, guilty feedback participants had more negative SC differential reactivity scores (appeared more deceptive) than guilty no-feedback participants. However, the difference between innocent feedback participants and innocent no-feedback participants that was found for chart 3 attenuated in charts 4 and 5. Thus, for charts 4 and 5, the effect of SC feedback in detecting deception in guilty participants remained; but by the fifth chart, the beneficial effects seen in the previous analysis for innocent participants was not statistically significant.

Composite APF also affected cardiograph responses, $\underline{F}(2, 272) = 3.25$, $\underline{p} < .05$, $\eta^2 = .02$. However, in this case the effects were relatively small and not consistently beneficial. Examination of the right panel of Figure 3 reveals that there was greater discrimination between guilty and innocent participants on the second chart for participants who received no feedback (circles) than for participants who received APF (triangles). In contrast, discrimination between guilty and innocent participants was greater on the third chart for those who received AFP than for those who did not. Indeed, innocent participants who received no feedback (open circles) evidenced slightly more negative cardiograph scores than did guilty participants who received no feedback (closed circles). This same trend remained for analyses conducted on charts 4 and 5.

There were no statistically significant effects of SC-feedback on habituation rates of respiration or cardiograph responses, nor were there effects of composite-feedback on the habituation rates of respiration responses.

Waveform Analysis

Diagnoses of truth and deception by the computer and by polygraph examiners are often based on increases in electrodermal and cardiovascular activity and respiration suppression. In the presence of APF, measures other than SC amplitude, cardiograph amplitude, and respiration excursion may be more diagnostic of truth and deception. To explore this possibility, split-plot ANOVA was used to test for differences in the shapes of various physiological responses to probable-lie and relevant questions over the 20-second interval that followed question onset. ANOVA was performed separately for SC, cardiograph, thoracic and abdominal respiration excursion, finger pulse amplitude, heart rate, and vagal tone. Between-group factors were Guilt (guilty and innocent) and Feedback (no-feedback, SC-feedback, and composite-feedback). Within-participant factors were Question Type (comparison and relevant) and Time (e.g., seconds).

Twenty second-by-second measurements were analyzed for all physiological measures except vagal tone (Podlesny & Raskin, 1978). Vagal tone was measured for each of four successive 5-second intervals. Of interest were Guilt X Question Type X Feedback, and Guilt X Question Type X Feedback X Time interactions. If the differences between comparison and relevant questions for guilty and innocent participants do not depend on the presence or type of APF, then measures found useful for individuals who do not receive APF also should be useful for individuals who do receive APF.

The Guilt X Question Type X Feedback interaction was statistically significant for thoracic respiration excursion, $\underline{F}(2, 203) = 3.37$, $\underline{p} < .05$, $\eta^2 = .03$. The means for comparison and relevant questions are presented in Figure 4. Baseline respiration measurements for neutral questions are included in Figure 4, but they were not included in the ANOVA.

As predicted, innocent participants generally evidenced less respiration activity in response to comparison questions than to relevant questions, whereas guilty participants showed less respiration activity in response to relevant questions. Figure 4 suggests that the interaction was due to an atypical pattern of responses by innocent participants who received SC feedback. As can be seen, the difference between comparison and relevant questions for innocent participants who received SC feedback was less than the difference for the participants in other groups.

A statistically significant effect was also found for the Guilt X Question Type X Feedback X Time interaction for finger pulse amplitude (FPA), $F(38, 3876) = 2.53$, $p < .02$, $\eta^2 = .02$. Figure 5 displays the FPA curves for the three feedback conditions. In general, a strong vasomotor response was indicated by a large reduction in the amplitude of finger pulses. The results given in Figure 5 indicate that guilty participants responded as predicted; they evidenced stronger vasomotor responses to relevant questions than to probable-lie questions across all feedback conditions. In contrast, innocent participants in the no-feedback and composite-feedback conditions showed little difference in their vasomotor responses to probable-lie and relevant questions. Only innocent participants in the SC-feedback condition responded more strongly to comparison questions than to relevant questions.

Discussion

The goals of this study were, through the use of APF, to attempt to increase the reliability and validity of the physiological measures obtained during a conventional CQT polygraph test, and to reduce habituation to repeated exposures to polygraph questions. Although not all of our hypotheses were substantiated, the results of the study that did confirm our hypotheses offer useful information to those conducting polygraph tests in the field.

Our first hypothesis, that the use of APF would increase the reliability of physiological measures was not statistically significantly substantiated. Although the use of both composite and skin conductance auditory feedback did increase the magnitude of the coefficient alpha index by five percentage points relative to the no-feedback condition, this increase in reliability probably is not "clinically" statistically significant, in that noticeable improvements in decision accuracy by reducing random variation in the way participants processed questions probably would not result from using APF. The data in Table 1, Figure 1, Figure 2 and Table 3, suggest the difference in percent correct decisions for both innocent and guilty participants in the SC Feedback condition was improved, but not statistically significantly so.

Although results suggest that APF may not enhance the reliability of CQT polygraph tests, an aspect of the protocol implemented in this study may account for the null effects observed from these data. Specifically, it may be the case that the lack of time elapsed from the "mock" crime committed by participants to the actual polygraph test, or the reward offered for an innocent verdict, caused the participants to experience enhanced intrinsic motivation to "perform" well on test and receive the cash bonus. Such motivation to be classified as innocent may not differ substantially from a suspect in a criminal investigation; however, rarely is an individual offered cash in exchange for an innocent verdict or given a polygraph test concerning alleged criminal involvement immediately following the actual crime.

Because the guilty participant was given the test immediately after committing the theft, his/her memory of the crime, and involvement with that criminal activity was probability more easily accessible affectively and difficult to suppress physiologically than the criminal who committed a theft in the days or even weeks preceding the polygraph test. Hence the effectiveness of APF might have been confounded by the degree of involvement of the participants with the recency of the crime, such that added benefits of detection over the No-feedback condition were lost.

Perhaps future investigations of APF on detection should include a mock crime that is committed three or more days preceding the polygraph investigation to more adequately represent the typical field polygraph investigation into alleged criminal activity. Using a smaller cash bonus may also serve to reduce the "ceiling effect" observed in this study. This idea, that the more "ego-involving" and relevant the participants perceive the testing situation to be, the less effective is the use APF was first mentioned in the Timm (1987) study of biofeedback effectiveness in assessing guilt as measured from the Guilty Knowledge Test (GKT); and the results of the first experiment of relevant (SS#) and non-relevant (geometric figures) items observed in the Stern, et al. (1981) study support such a claim. Thus, these studies, coupled with the results found in this investigation, indicate that further investigation into perceived participant involvement with the test is necessary.

Our hypothesis that APF would decrease habituation rates as participants completed successive polygraph charts was substantiated. Specifically, SC amplitude for those in the SC Feedback condition did not evidence the decrease

typically seen as suspects complete the second and third polygraph charts. In fact, by the completion of the third chart, there still existed a greater delineation between innocent and guilty participants who received SC feedback or composite feedback than those who received no feedback. This effect persisted even after including two additional charts of data for the guilty participants. These two charts are not part of a standard polygraph test and further support the strength of APF in detecting guilty participants who may be required to complete a longer version of a standard polygraph test. Thus, APF may serve to decrease fatigue effects or lack of involvement in the test commonly observed after repeated exposures to the test questions for guilty participants and increase the usefulness of latter charts for detecting deception.

Because diagnoses of truth and deception by polygraph examiners are often based on increases in electrodermal and cardiovascular activity and respiratory suppression, another goal of this investigation was to attempt to examine alternative methods of interpreting physiological responses in the presence of APF that may be more diagnostic of truth or deception. As predicted, investigations into second-by-second measurements of thoracic respiration excisions for innocent participants showed that their responses were more suppressed to comparison questions than to relevant questions, whereas guilty participants evidenced more suppression to relevant questions than to comparison questions.

A statistically significant effect of finger pulse amplitude was also found for guilty participants in all feedback conditions. As expected, a stronger vasomotor response was observed in guilty participants for relevant questions than comparison

questions. Results for innocent participants suggest that SC Feedback enhances the predicted increase in vasomotor responding to comparison questions. These promising results obtained for SC Feedback are also consistent with the findings of the Stern, et al. (1981) study that found greater success with SC Feedback than the heart rate feedback.

Overall, the results suggest a number of implications concerning the use of APF during CQT polygraph tests. Although detection rates did not appear to be enhanced by APF in this study, further investigation into the benefits of APF are needed. The use of APF in this investigation was shown to offer at least one clear benefit for the CQT polygraph test. APF decreases the rate of habituation over repetitions of the question sequence and allows for greater discrimination between innocent and guilty participants as the polygraph test progresses.

References

Ben-Shakhar, G. (1991). Clinical judgment and decision-making in CQT-polygraphy: A comparison with other pseudoscientific applications in psychology. Integrative Physiological and Behavioral Science, 26, 232-240.

Bradley, M. T. & Janisse, M. P. (1981). Accuracy demonstrations, threat, and the detection of deception: Cardiovascular, electrodermal and pupillary measures. Psychophysiology, 18, 307-314.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, N. J.: Prentice Hall.

Keppel, G. (1991). Design and analysis: A researcher's handbook (3rd Ed.). Engelwood Cliffs, NJ: Prentice Hall.

Kircher, J. C. (1981). Psychological and psychophysiological processes underlying the detection of deception. Unpublished manuscript. Department of Psychology, University of Utah, Salt Lake City, UT

Kircher, J. C. & Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. Journal of Applied Psychology, 73, 291-302.

Kircher, J. C. & Raskin, D. C. (1994). The computerized polygraph system. Version 2.00 Manual. Salt Lake City, UT: Scientific Assessment Technologies.

Kircher, J. C., Woltz, D. J., Bell, B. G., & Bernhardt, P. C. (1998). Effects of audiovisual presentation of test questions during relevant-irrelevant polygraph examinations and new measures. Final report to the U. S. Government. University of Utah, Salt Lake City, UT

Podlesny, J. A., & Raskin, D. C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. Psychophysiology, 15, 344-358.

Porges, S. W., Bohrer, R. E., Cheung, M. N., Drasgow, F., McCabe, P. M., & Keren, G. (1980). New time series statistic for detecting rhythmic co-occurrence in the frequency domain: The weighted coherence and its application to psychophysiological research. Psychological Bulletin, 88, 580-587.

Raskin, D. C., Kircher, J. C., Honts, C. R , & Horowitz, S. W. (1988). A study of the validity of polygraph examinations in criminal investigation (Grant No. 85-IJ-CX0040). Salt Lake City: University of Utah, Department of Psychology.

Raskin, D. C., Honts, C. R., & Kircher, J.C. (1997). Polygraph techniques: Theory, research and applications from the perspective of scientists-practitioners. In D. L. Faigman, D. Kaye, M. J. Saks, & J. Sanders (Eds.), Scientific evidence reference manual (pp.551-582). St. Paul, MN :West Publishing Co.

Reid, J. E., & Inbau, F.E. (1977). Truth and deception: The polygraph ("Lie Detector") technique. Baltimore: Williams & Wilkins.

Stern, R.M., Breen, J. P., Watanabe, T., & Perry, B.S. (1981). Effects of feedback of physiological information on responses to innocent associations and guilty knowledge. Journal of Applied Psychology, 66, 677-681.

Timm, H.W. (1987). Effect of biofeedback on the detection of deception. Journal of Forensic Sciences, 32, 736-746.

Table 1. Percent computer outcomes and percent correct decisions excluding inconclusive outcomes (n = 35 per group)

		Correct	Wrong	Inconclusive	Correct Decisions
No Feedback	Innocent	69	23	9	75
	Guilty	74	14	11	84
SC Feedback	Innocent	74	17	9	81
	Guilty	77	17	6	82
Composite Feedback	Innocent	69	17	14	80
	Guilty	60	23	17	72

Table 2. Coefficient alphas (internal consistency reliability) for computer indices of differential reactivity across five polygraph charts

Physiological Measure	No Feedback (n = 61)	Skin Conductance Feedback (n = 61)	Composite Feedback (n = 60)
Skin Conductance Amplitude	.85	.91	.80
Cardiograph Amplitude	.40	.53	.71
Respiration Excursion	.81	.79	.71
Mean	.67	.74	.74

Figure 1. Skin conductance, cardiograph, and respiration indices of differential reactivity for no-feedback and SC feedback groups

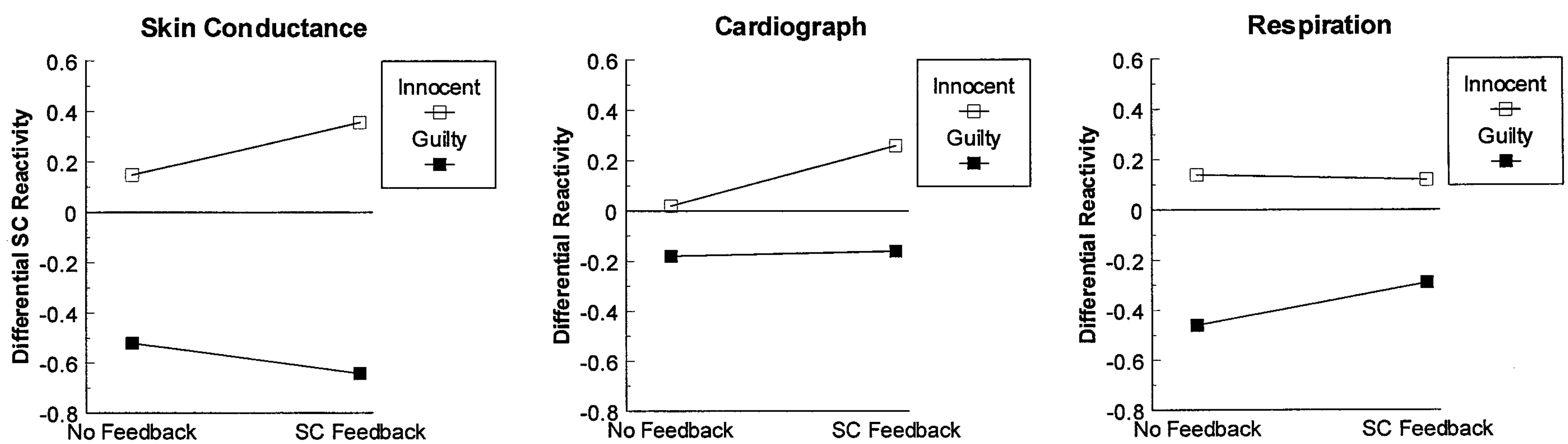

Figure 2. Skin conductance, cardiograph, and respiration indices of differential reactivity for no-feedback and composite feedback groups

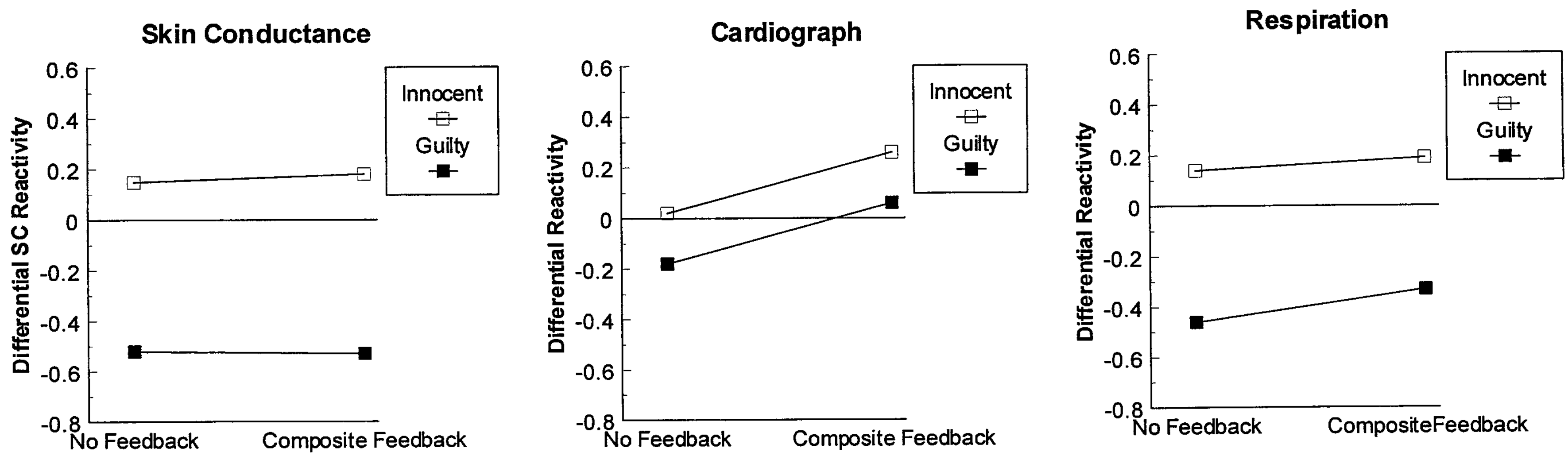

Table 3. Percent dichotomous computer outcomes based on first three charts (n = 35 per group)

		Correct	Wrong
No	Innocent	71	29
Feedback	Guilty	83	17
SC	Innocent	74	26
Feedback	Guilty	83	17
Composite	Innocent	77	23
Feedback	Guilty	71	29

Figure 3. Effects of feedback on habituation of skin conductance and cardiograph responses

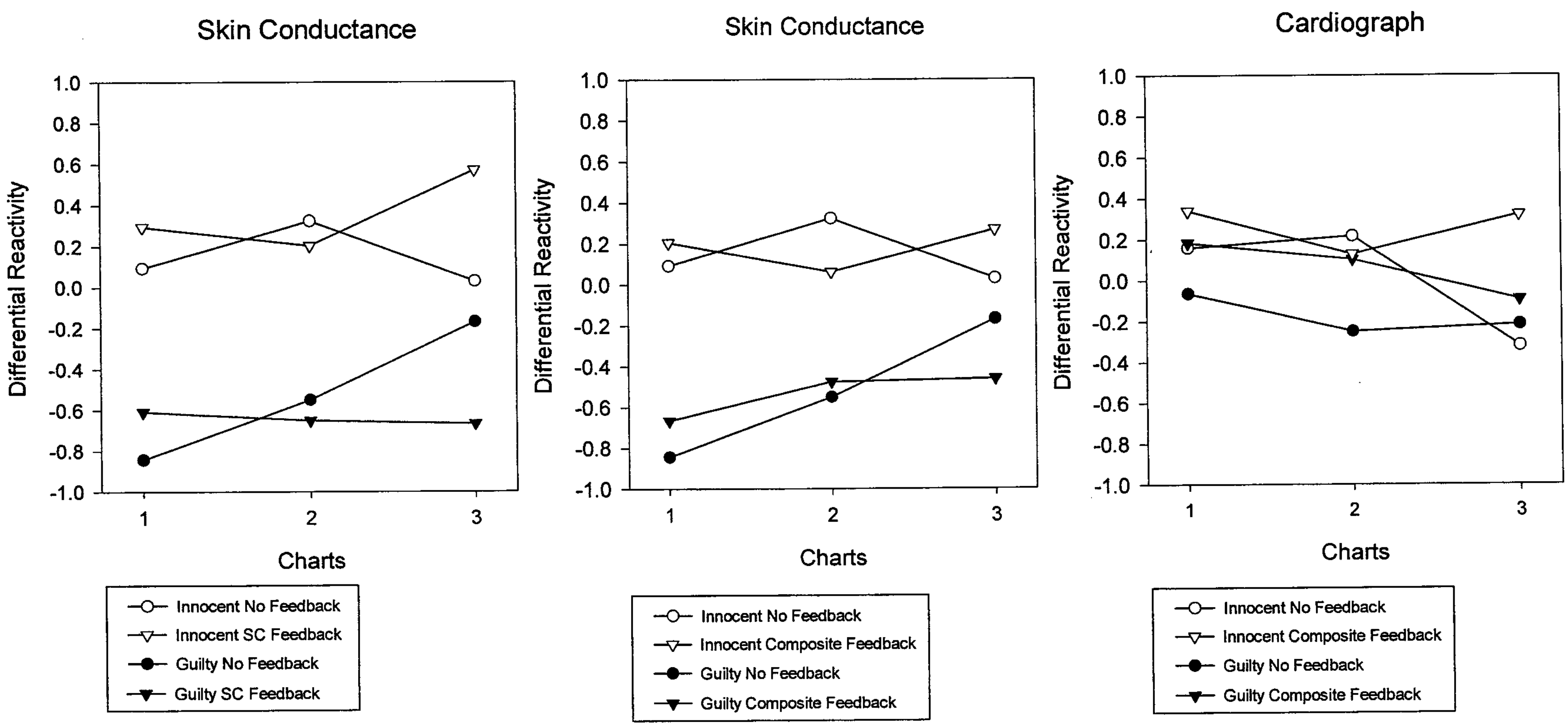

Figure 4. Thoracic respiration excursion for neutral, comparison, and relevant questions under no-feedback, SC feedback, and composite

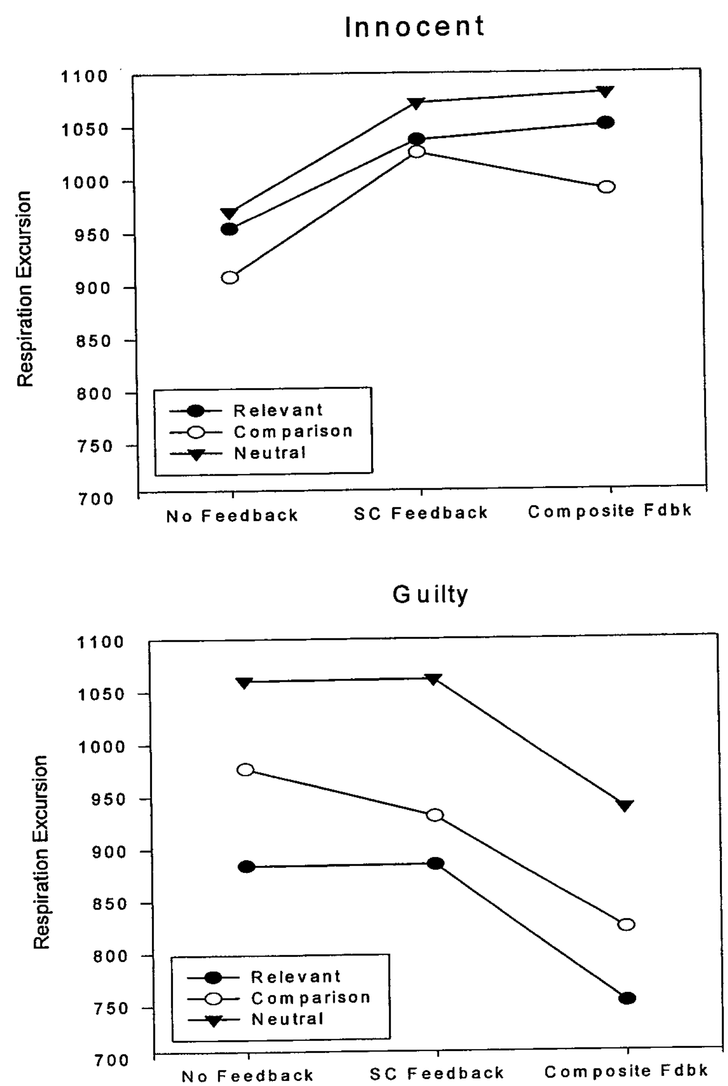

Figure 5. Finger pulse amplitude responses to comparison and relevant questions for guilty and innocent participants in three feedback conditions